Looking Inside Non-coding Chloroplast Regions of *Calophyllum brasiliense* (Calophyllaceae) to Understand Its Southernmost Population Distribution

Cecilia Beatriz Percuoco[1], Liliana Noelia Talavera Stéfani[1], Manuela Edith Rodríguez[2], Naiké Lucía González[3], Juan Fernando Crivello[4], Jorge Víctor Crisci[5], Carina Francisca Argüelles[1, *]

[1]Laboratorio GIGA, Instituto de Biología Subtropical, nodo Posadas, Universidad Nacional de Misiones, Consejo Nacional de Investigaciones Científicas y Técnicas. Posadas, Argentina

[2]Cátedra de Biología Vegetal. Facultad de Ciencias Exactas Químicas y Naturales, Universidad Nacional de Misiones. Posadas, Argentina

[3]Cátedra de Sistemática Teórica. Facultad de Ciencias Exactas Químicas y Naturales, Universidad Nacional de Misiones. Posadas, Argentina

[4]Cátedra de Introducción a las Ciencias de la Tierra. Facultad de Ciencias Exactas Químicas y Naturales, Universidad Nacional de Misiones, Posadas, Argentina

[5]Laboratorio de Sistemática y Biología Evolutiva. Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata. La Plata, Argentina

Email address:
franciscarguelles@fceqyn.unam.edu.ar (C. F. Argüelles)

Abstract: In recent years the growing interest in the conservation of Paraná River's riparian forest led to the discovery of botanical novelties for Argentina. Populations of *Calophyllum brasiliense* Camb. (Calophyllaceae), a typically flooded lowlands species, were identified in the remaining hygrophile forest of northeast Argentina and southeast Paraguay. Deforestation and flooding, due to the construction of dams, have caused these populations to suffer intensive fragmentation. The aim of this work was to infer phylogeographic relationships among five populations of *C. brasiliense*, three from Argentina and two from Paraguay, which represent the southernmost points of species' distribution. We also compared them with samples of a *C. brasiliense* population from Mexico, the northernmost edge of the species distribution. The chloroplast intergenic spacers *petG-trnP*, *psbJ-petA* and the *trnL*-UAA chloroplast intron were amplified from leaves' DNA. A total of 2234 bp were characterized once the three regions were analyzed. The three chloroplast regions showed nucleotide differences, represented by InDels, inversions and a few SNPs; however, only the *trnL* intron was selected for further phylogeographic analysis due to the amount of the information obtained for all populations. Based on *trnL* intron, it was possible to estimate nucleotide and haplotype diversity ($\pi = 0.00237$ and $Hd = 0.29600$, respectively). Three haplotypes were identified, which allowed Argentinean, Paraguayan and Mexican populations to be differentiated. Based on the three haplotypes found, we discuss and propose a model for a *C. brasiliense'* geographic dispersion and historical colonization routes, including an alternative new one to the well-known of the Paraná River.

Keywords: *C. brasiliense*, cpDNA, *petG-trnP*, *psbJ-petA*, *trnL* Intron

1. Introduction

The Paranaense and the Amazonic Biogeographic Subregions are separated by the Chaqueña Subregion, all three cover a great portion of the South America and belong to the Neotropical Region [1] (Figure 1a). The Chaqueña Subregion is known as the major South American disjunction because it is mainly represented by savanna species, with sparse and dry vegetation, which acts as a dispersion barrier between the Paranaense and the Amazonic rich tropical forest [2, 3].

Within the Paranaense Subregion, a unique forest formation stands out for growing in almost permanent flooded soils. This forest type shows a differentiated structure

and floristic composition, which is known as "floresta higrófila", "mata de brejo" [4] or "seasonal semideciduous forest" with permanent fluvial influence [5]. There are typical or exclusive species at hygrophile forests, as is the case of *Calophyllum brasiliense* Camb., which are absent in other forest types [6].

C. brasiliense (Calophyllaceae) exhibits a broad geographic distribution which contrasts with high habitat specificity, since it exclusively grows in areas saturated with water [5, 7]. This ability provides advantages in comparison with other species, which explains its highly concentrated occurrence in damper regions [8-11]. The distribution of *C. brasiliense* extends from Mexico to northeast Argentina [12, 13] and is commonly known as "guanandí" or "cedro de pantano" in Brazil and "arary" in Argentina and Paraguay, among other names. Its economic importance is due to its high quality wood and several bioactive components that are of medical interest [12]. The oil from the seeds has been successfully used to cure some skin diseases [14] and extracts from different plant parts exhibit antiviral activity by inhibiting HIV-1 [15-17]. In the same way, stem bark extracts have been demonstrated to have anti-leukemic effects [18, 19] and a lethal action against the parasitic protozoan *Tripanozoma cruzi* [20]. Some bioactive compounds isolated from its leaves have proven to have a significant molluscicidal activity against *Biomphalaria glabrata* [14] while the antibacterial potential against *Mycobacterium tuberculosis* has been recently reported [21].

The Paranaense Subregion has faced threats for decades, which have transformed it into one of the most vulnerable forests in the world with a current rate of change that seems to be accelerated. Most of the original forest has disappeared or diminished, especially in southeastern Brazil, eastern Paraguay and northeastern Argentina [22, 23]. While the loss of biodiversity begins with the reduction of ecological interactions and genetic variability, it ends with the local extinction of plant and animal populations [24].

C. brasiliense populations have been identified in the southern zone of the Paranaense Forest, some of them located in northeastern Argentina and a few more in southeastern Paraguay. The populations are in riverine forest, both sides of the Paraná River upstream Yacyretá dam and in Yacyretá Island (Paraguay). The dynamics and structure of these populations were modified (Cardozo et al., *unpublished data*) following the last increase in river level associated with dam requirements. The last identified population is nearby the Iberá wetlands natural Reserve, in the Department of Ituzaingó (Corrientes - Argentina). While this small patch has not been affected by increasing water levels, it has been impacted by surrounding forest of *Pinus* and *Eucalyptus*.

As a component of biodiversity conservation, the underlying genetic diversity within these populations would be useful in the identification of conservation units as evolutionary significant units (ESUs) or management units (MUs), both fundamental for proper short and long-term management [25]. Currently only a rapid genetic characterization for the first two *C. brasiliense* populations

identified in Argentina exists [26]; therefore, we would like to expand our understanding to include the origin and evolution of these populations.

Since the first pioneer studies that use chloroplast DNA (cpDNA) variability to solve phylogenetic questions were published [27-30], many research groups have focused in the chloroplast polymorphic regions due to the well-known, advantageous characteristics of haploid genomes. The small size and largely or completely maternal inheritance (except conifers) [31], as well as the typical absence of recombination and the structural conservative evolution, are properties that make cpDNA an ideal tool for evolutionary studies with species and populations [32-34]. In the last few years as sequencing technology has increased in availability, the taxonomic levels analyzed with cpDNA have decreased from families to species and populations [35-38].

In the present work we hypothesize that the southernmost *C. brasiliense* populations historically belonged to a continuum of large forest that underwent a profound fragmentation process that resulted in genetic differentiation among the remaining populations. Based on this hypothesis, the aim of this work was to: 1) evaluate three cpDNA regions in the Argentinean and Paraguayan populations of *C. brasiliense* in order to understand their genetic and phylogeographic characterization, and 2) contrast these populations with individuals from the northernmost point in the species' distribution.

2. Materials and Methods

2.1. Study Area

In the present work, three Argentinean and two Paraguayan *C. brasiliense* populations were sampled. Misiones and Corrientes are northeastern Argentinean provinces bordered by the Paraná and Uruguay rivers (Figure 1c, Table 1). Misiones contains the largest remaining tract of Paranaense Forest [23]. Both provinces are characterized by humid, subtropical climate with no distinct dry season [39]. Two of the Argentinean populations are from Corrientes [Puerto Valle (PV) and Rincón Ombú (RO)] and the third one is located in Misiones [Department of San Ignacio (SI)]. In this region, the Paraná River represents the boundary line between Argentina and Paraguay. Inside Paraguay, *C. brasiliense* populations studied were both located in the Department of Misiones, in Yacyretá Island (IY) and Ayolas city (AY) (Figure 1c, Table 1). In addition to these samples, a set of Mexican (MX) samples representing the species' northernmost distribution from Jaguaroundi Ecological Reserve, Coatzacoalcos city, Veracruz State was included (Figure 1b, Table 1).

2.2. Sample Collection

In each population, 2 – 10 individuals were considered (Table 1). Four to five young leaves from adult trees from each population were collected and after cleaning with 70% ethanol were stored dry in individually labeled Ziploc bag

containing silica gel. They were maintained at room temperature until DNA extraction. The unique criterion for choosing each individual tree was a minimal distance of 10 m among them.

2.3. DNA Extraction

Completely dried leaves were ground to a fine powder in a ceramic pestle using liquid nitrogen. The total genomic DNA was isolated from leaf tissue (120 mg) based on the CTAB protocol [40] with modifications which briefly include the incorporation of 2% polyvinylpyrrolidone (PVP), 5 mM ascorbic acid, 4 mM sodium diethyldithiocarbamatetrihydrate (DIECA) and 1.2% β-mercaptoethanol into the digestion buffer in a 2 ml plastic tube to improve the homogenization at 37°C for 3 h [26]. After chloroform: isoamyl alcohol (24:1) organic extractions, the DNA was precipitated with 2/3 isopropanol at -20°C for 1 h. The DNA isolation was verified and semiquantified electrophoretically in 1% agarose gels, stained with 0.5µg/ml ethidium bromide against a mass ladder (BioRad), yielding 10 - 50 ng/µl.
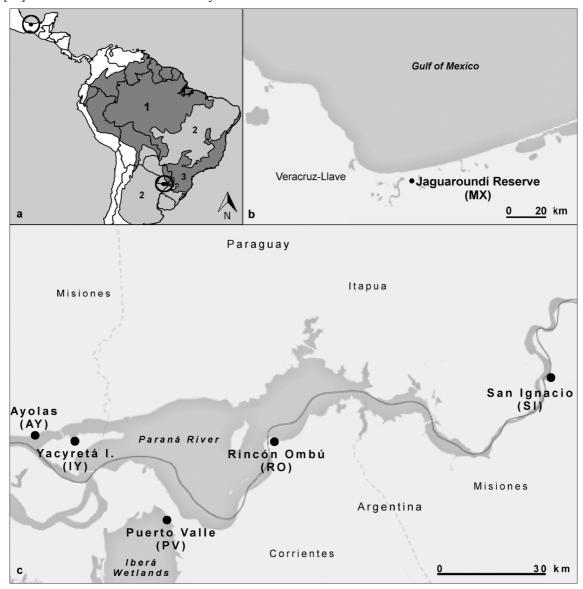

Figure 1. Location of the five studied populations within the biogeographic regions (modified from Morrone, 2001). a. Biogeographic Subregions cited along the text: 1. Amazonic Subregion. 2. Chaqueña Subregion. 3. Paranaense Forest Subregion. Circles: Sampled populations (POP). b. Mexican C. brasiliense POP. c. Argentinean and Paraguayan C. brasiliense POP.

2.4. Primer Selection

According to the conserved nature of chloroplast in higher plants [41, 42] primers designed for other angiosperms species were tested [43, 35]. Two chloroplastic intergenic spacers, *petG-trnP* (petG: GGT CTA ATT CCT ATA ACT TTG GC; trnP: GGG ATG TGG CGC AGC TTG G) and *psbJ-petA* (psbJ: CTC TTT GGT TGA TAG GTA CTG; petA: GGA GAT GCA GAG ATA GTA C), and the *trnL*-UAA intron were amplified (*trnL*c(F): CGA AAT CGG TAG ACG CTA CG; *trnL*d(R): GGG GAT AGA GGG ACT TGA AC). All oligonucleotides were synthesized by IDT® or Macrogen Inc.

2.5. Amplification Reactions

The reactions were optimized with 2 µl of 1/50 DNA dilutions (25 ng/µl), 2 mM MgCl$_2$, 0.2 mM of dNTPs Mix, 5 µg/µl of BSA, 0.1 µM of each primer, and 1U of Taq polymerase (Fermentas) with 1X SO$_4$(NH$_4$)$_2$Taq Buffer, in a 25 µl final volume. For *psbJ-petA* and *petG-trnP* the PCR profile consisted of a 5 minutes denaturation at 80°C, followed by 35 cycles at 95°C for 1 minute, 50°C for 1 minute and 65 °C for 4 minutes [36]. For the *trnL* intron the PCR profile consisted of an initial 5 minutes step at 94°C, followed by 35 cycles of 93°C for 1 minute, 62°C for 1 minute and 72°C for 2 minutes. A final elongation segment of 4 minutes at 72 °C was necessary. For the *trnL* intron an internal primer pair was designed using Primer3 software [44, 45] and tested with Primer-BLAST (Cbra_F: 5'AAAATGGGCAATCCTGAGC, Cbra_R: ATGGGACTCTATCTTTATTCTCG). PCR profile matched that used for intron *trnL* but without the initial denaturation step. All amplifications were carried out on a Perkin Elmer GeneAmp System 9600 thermal cycler (Applied Biosystems). To minimize the potential for contamination in all reactions negative controls (no DNA added) were included. Fragment sizes were evaluated electrophoretically in agarose gel (1.5%) using 1X TBE buffer and BrEt (0.5 µg/ml) against 100 bp ladder (Productos Bio-Lógicos®) as molecular marker. Gels were run at 5 V/cm and visualized under UV light [46].

Table 1. Details of the analyzed populations (POP) and sample sizes (N). ARG: Argentina. PY: Paraguay. MX: Mexico.

POP	Locality	Geographic reference	N
San Ignacio (SI)	San Ignacio (Misiones - ARG)	27º16'34.4"S - 55º34'11.9"W	11
Rincón Ombú (RO)	Ituzaingó (Corrientes - ARG)	27º24'54.42"S - 56º29'43"W	11
Puerto Valle (PV)	Ituzaingó (Corrientes - ARG)	27º35'23.84"S - 56º31'34.17"W	10
Yacyretá Island (IY)	Misiones (PY)	27º24'53.88"S - 56º45'21.25"W	3
Ayolas (AY)	Misiones (Misiones - PY)	27º24'6.48"S - 56º51'17.75"W	2
Jaguaroundi Reserve (MX)	Coatzacoalcos (Veracruz Llave- MX)	18°7'20.27"N – 94º21'45.15"W	6

2.6. Sequencing Reactions and Analysis

Amplicons of individuals randomly selected from each population were sequenced by Macrogen Inc. using forward and reverse primers. Products were previously purified using GFX PCR DNA #28-9034-70 (GE). The three amplified regions were sequenced for 2-11 individuals of each population. For each analyzed individual, forward and reverse reads were edited and a consensus sequence was created. Three separated alignments were done to identify polymorphic sites and to describe the observed haplotypes for each analyzed region. Additional alignments were carried out among haplotypes found in *C. brasiliense* and the most identical GenBank sequences using the Basic Local Alignment Search Tool BLASTn 2.2.25 [47]: *Hevea brasiliensis* (*psbJ-petA* and *petG-trnP*) and *Calophyllum inophyllum (trnL* intron). All consensus sequences were aligned for SNPs and InDels identification. All sequences edition and alignment were carried out using Bioedit v.7.0.9.0 [48] with ClustalW setting default parameters.

2.7. Phylogeographic Analysis

Based on the quantity and quality of sequences obtained per population, *trnL* intron was selected for the phylogeographic analysis. The alignment analysis allowed estimation of haplotype number (*h*), gaps, polymorphic sites (*S*), allele number per polymorphic site, parsimonious or informative sites, and singletons (non-informative sites). Haplotype diversity (*Hd*) and nucleotide diversity (π) were computed [49]. The inter-population differentiation was assessed through F_{ST} statistic. Tajima's *D* [50] and Fu's F_S [51] were estimated to provide inference on sequence neutrality and to evaluate past departures in population size with a confidence of 99% (p<0.01). All parameters estimations were performed using DnaSP v.5.10.1 [52].

Phylogeographic relationships among *C. brasiliense* populations were determined using haplotype networks constructed with the median-joining algorithm in Network v.4.611 [53]. The evolutionary divergences among haplotypes were illustrated by means of a phylogenetic tree reconstructed under maximum parsimony criteria using MEGA v.5.05 [54]. A *trnL* intron sequence from *C. inophyllum* (AB817676.1) was included as outgroup to set the tree root and give directionality to the observed changes.

3. Results

We successfully amplified the three chloroplastic regions in a variety of individuals. Furthermore for *trnL* intron the specific primers designed (Cbra_F and Cbra_R) allowed us to obtain good quality reads for all 43 individuals analyzed. All populations were represented by two or more sequences of the three chloroplastic regions, although *psbJ-petA* and *petG-trnP* intergenic flanking primers failed to amplify for some individuals. As the total number of sequences was different for each region, they were analyzed independently. The three regions added up 2234 bp.

psbJ-petA: A total of 30 sequences of 1200 bp in length were aligned, which enabled us to recognize two haplotypes (*H1_JA*, KJ093668 and *H2_JA*, KJ093669) that were differentiated by two InDels, a substitution and a T$_n$ mononucleotide chloroplastic microsatellite (cpSSR) (Table 2a). *H1_JA* was exclusively observed in individuals belonging to Argentina and Paraguay populations whereas *H2_JA* was shared among the six Mexican samples analyzed.

Table 2. *Alignments showing variable positions within three cpDNA non-coding regions analyzed in C. brasiliense compared with sequences from Hevea brasiliensis and Calophyllum inophyllum. a. psbJ-petA. b. petG-trnP c. trnL intron. ARG: Argentina. PY: Paraguay. MX: Mexico.*

a. psbJ-petA

Haplotype	Positions	POP	# GenBank
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		
	0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1		
	9 1 1 1 1 1 1 1 1 1 2 2 2 2 2 0 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 4		
	8 0 1 2 3 4 5 6 7 8 9 0 1 2 3 6 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 2		
H1_JA	T A T T A T T A T A T T A T T A T T A T T A T A T A A A T T C G A A T A A C	ARG, PY	KJ093668
H2_JA	- - - - - - - - - - - - - - - - G - T	MX	KJ093669
H. brasiliensis	N N N N N N N N N N N N N N · T C · · · · T · G · C · C · · - - - - - - · · ·	-	HQ285842

b. petG-trnP   **c. trnL intron**

Haplotype	Positions	POP	# GenBank	Haplotype	Positions	POP	# GenBank
	0 0 0 0 0 0 0 0 1 1				2 2 3 3 3 3 3 3 3 3 3 4		
	9 9 9 9 9 9 9 9 0 0				7 9 1 1 3 3 3 3 4 4 8 1		
	2 3 4 5 6 7 8 9 0 1				5 3 6 7 5 6 7 8 9 0 1 1 6		
H1_GP	C C T T T C - - - - - - C C G T	ARG, PY	KJ093665	H1_Lint	G G T C - - - - - - - T T	ARG, PY	KF854307
H2_GP	A C G G G A - - - - - - A A G G	MX	KJ093666	H2_Lint	G G G A - - - - - - - T T	IY	KF854308
H. brasiliensis	· · · · · · T T T A A T · · · C	-	HQ285842	H3_Lint	G T G A T A C A A A T G	MX	KF854309
-	- - - - - - - - - - - - - - - - -	-		*C. inophyllum*	T T G A T A C A A A C ·	-	AB817676

N: any nucleotide; - : gap; · : identity

petG-trnP: A total of 19 sequences of 449 bp in length were aligned, which again enabled us to recognize two haplotypes (*H1_GP*, KJ093665 and *H2_GP*, KJ093666) that were differentiated by a unique mutational step consistent in a 8 bp inversion flanked by two 10 bp inverted repeats in both directions (Table 2b). As in the previous region, *H1_GP* was unique for Argentinean and Paraguayan *C. brasiliense* populations whereas *H2_GP* was observed in all six samples from Mexico.

trnL intron: A total of 43 sequences of 450 bp in length were aligned, which enabled us to recognized three haplotypes (*H1_Lint*, KF854307; *H2_Lint*, KF854308; *H3_Lint*, KF854309) that were differentiated by a 7 bp InDel and 4 informative SNPs (293, 316, 317, 416) with two variants, but no singletons were observed (Table 2c). Two of these positions were adjacent (316, 317) which seems to be the result of an inversion that could involve a unique mutational step. On the other hand, the InDel consisted of a 7 bp duplication from 335 to 341 nt. *H1_Lint* was present in all Argentinean and Paraguayan populations, *H2_Lint* was found in the Ayolas population whereas *H3_Lint* was common to all Mexican individuals.

The total haplotype diversity (*Hd*) and the total nucleotide diversity was 0.28600 and 0.00237, respectively; however, values distinct to zero were obtained for the Yacyretá Island population (Table 3). The estimated F_{ST} indexes did not differentiate the three Argentinean populations (SI, RO, PV) or the two Paraguayan ones (IY, AY); however, high estimates were obtained between the Paraguayan and the Mexican populations (F_{ST} = 0.750, p<0.01) and between the Argentinean and the Mexican populations (F_{ST} = 1.000, p<0.01) (Table 4).

The neutrality tests were only carried out on Paraguayan populations because they exhibited haplotype diversity greater than zero. Both Tajima's and Fu's test were not significant at p<0.01 (Table 5), indicating that the variants found within them are in accordance with the neutral evolution model and do not reflect past demographic events. The evolutionary relationships among the three identified haplotypes based on *trnL* intron are represented in the phylogenetic tree shown in Figure 2, where the *C. inophyllum trnL* intron sequence (#AB817676.1) was included. The haplotype frequencies and distribution along all populations are summarized in the haplotype network (Figure 3) where the Paraguayan populations were grouped together.

4. Discussion

The use of non-coding chloroplast regions for phylogeographic inferences has proven to be useful in several population studies; however, there is agreement in recognizing that each region has its own characteristics differing across taxonomic groups [36, 55, 56]. Therefore when choosing a region suitable for intraspecific analyses it is important to take into account that chloroplast genome is extremely conserved within species, at least in the most widely used intergenic spacers. However, there are other more variable unexplored regions that are rarely employed [37].

The three chloroplastic regions analyzed in the present work were able to highlight haplotype variants among *C. brasiliense* populations. The intergenic regions *psbJ-petA* and *petG-trnP* have not been frequently used in phylogeographic analyses; however we could identify InDels in both regions for the analyzed populations adding a nucleotide substitution and a mononucleotide microsatellite in the case of *psbJ-petA*. These two spacers allowed us to distinguish the southernmost Argentinean and Paraguayan populations from samples that represented the northernmost populations in *C. brasiliense*

distribution range. Among the studies that have used *petG-trnP*, a 350 bp intergenic spacer was sequenced in the Asian tropical tree *Trochodendron aralioides* and reported four substitutions analyzing four individuals from each of the 24 populations in Taiwan and Japan [35]. Similarly, two polymorphic sites and two InDels were detected in the 441 bp spacer while inferring phylogeographic relationships among 17 populations of *Cunninghamia konishii,* an Asian conifer [57].

Table 3. *Diversity parameters estimated from the alignment of 43 trnL intron sequences of C. brasiliense populations. S: segregant positions. H/N: haplotype/number of individuals. π: nucleotide diversity. Hd: haplotype diversity.*

POP	S	H/N	π	Hd
SI	0	H1/11	0.00000	0.00000
RO	0	H1/11	0.00000	0.00000
PV	0	H1/10	0.00000	0.00000
IY	2	H1/2. H2/1	0.03010	0.66700(0.314)
AY	0	H1/2	0.00000	0.00000
MX	0	H3/6	0.00000	0.00000
Total	4		0.00237	0.28600 (0.081)

Table 4. *F_{ST} differentiation indexes obtained among analyzed C. brasiliense populations. POP were clustered in ARG (SI, RO, PV), PY (IY, AY) and MX.*

POP	ARG	PY	MX
ARG	0.000	0.000	1.000
PY	0.000	0.000	0.750
MX	1.000	0.750	0.000

Table 5. *Neutrality test carried out from the 43 trnL intron sequences alignment of C. brasiliense individuals derived from the Paraguayan populations. CI: confidence interval.*

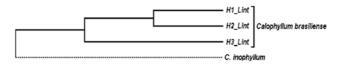| Neutrality test	Value	99% CI
Tajima's *D*	0.31826	-1.85076 – 2.83370
Fu's F_S	2.099	-4.84068 – 7.10335

Figure 2. *Phylogenetic tree based on the 43 trnL intron sequences aligned. C. inophyllum trnL intron sequence (#AB817676.1) was used to set the root of the tree. Terminal labels refer to haplotypes H1_Lint to H3_Lint found in Argentinean, Paraguayan and Mexican C. brasiliense populations.*

It is interesting to note that the type of mutations selected to reconstruct the phylogeographic history of a taxon indicate the importance of considering its particular characteristics and the possibility of them being homoplasies. For example, chloroplastic SSRs (cpSSRs) often arise independently in different lineages; therefore, several authors have excluded the frequent poliA/T cpSSRs considering them inappropriate for phylogeographic analysis at population level [35, 58-60]. Another type of variant extensively distributed along the majority of chloroplast genomes are shorts InDels (1-10 bp) corresponding to the 25-50% of mutations in non-coding

regions [39, 61]. This type of polymorphism was observed in the *psbJ-petA* spacer of *C. brasiliense* populations. On the other hand an 8 bp inversion flanked by inverted repeats was observed in *petG-trnP*, which probably made possible the formation of a stem-loop structure favoring intra-molecular recombination events. This kind of inversions has been reported for other intergenic and intronic regions [62-64] in many cases suggesting that they are common enough to occur independently provoking evolution parallelism [61]. However in the present study the difficulties to obtain amplicons from some individuals and the fact that no variants were found among the southernmost populations brought us to exclude *petG-trnP* and *psbJ-petA* from the following analysis.
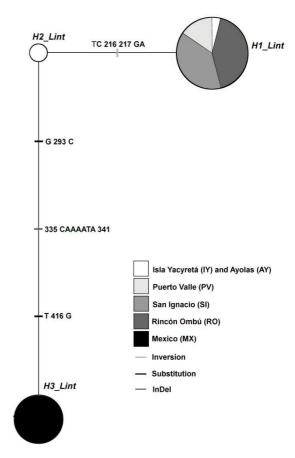

Figure 3. *Haplotype network based on C. brasiliense trnL intron from ARG (SI, RO and PV), PY (IY and AY) and MX populations.*

The plastid *trnL* intron is one of the most used in plant systematics and phylogeographic studies since the 1990's [65-67]. Some authors have suggested it as an appropriate region for DNA barcoding [68]. In contrast, others have found that resolution within this region is low [36], in part due to a highly conserved stem-loop structure, P6-loop, mixed with less constrained variable regions used by ecologist for studying highly degraded DNA [69]. Intraspecific variation was found in the phylogenetic study of *Clusia* genera where three mononucleotide and one dinucleotide microsatellites, all four within a 671 bp sequence of *trnL* intron, were reported [70]. In this study, we

observed only a mononucleotide repeat, whereas the three others were interrupted when comparing *C. brasiliense* with *Clusia* sp. sequences (Percuoco, 2014, *data not shown*). However, the polymorphisms found among the *C. brasiliense* populations analyzed in the current work support the use of *trnL* intron for population approaches in *C. brasiliense*. As reported, three haplotypes were found among the *C. brasiliense* populations studied and nucleotide diversity was due to four transversions and a 7 bp InDel. As was mentioned above, two of the four transversions were adjacent positions that could probably correspond to an inversion that was shared with the Mexican samples and one individual from Paraguayan Yacyretá Island. It is necessary to analyze a greater number of individuals in order to determine the phylogeographic significance of this inversion because we could not exclude that this mutation has arisen independently in these populations. A similar situation was reported in the analysis of *Trochodendron aralioides* populations, where an 11 bp inversion was observed in three individuals from two distinct populations [35]. These authors sampled four individuals per population and also suggested that analyzing more individuals was needed in order to discuss the significance of this variant.

Our results for the *trnL* intron region allowed us to differentiate the Argentinean and the Paraguayan populations, where two haplotypes were identified (*H1_Lint* and *H2_Lint*). A third haplotype (*H3_Lint*) was observed exclusively in the Mexican samples. With respect to the proposed hypothesis, these results seem to be in accordance with a common origin of the Argentinean and Paraguayan populations all sharing *H1_Lint*, which could indicate that the five analyzed populations conformed a forest continuum in the past. The *H2_Lint* found in one of the Paraguayan Yacyretá Island (IY) could be indicating an early divergence from Argentinean populations, assuming their common origin. *H1_Lint* found within all five populations evaluated in the present work demonstrated that the use of cpDNA genetic data reveal stronger genetic structure when compared with other molecular markers. Therefore this characteristic should be highly valued in the case of *C. brasiliense* phylogeography studies from other geographical regions. An average of 46 π estimates from phylogeographic studies that used cpDNA showed a mean of 0.00070 [71]. Later this value was doubled with the extended use of sequencing technology, and ranged from 0.00065 to 0.15000 [72].

In plant phylogeography studies it seems to be common to sample a few individuals from many populations, emphasizing inter- over intra-population differences [35, 61]. It has been demonstrated that in order to have more precise estimates is more convenient to sample few individuals from many populations than many individuals from a few populations [73]. Eight individuals are enough to obtain reliable estimates, providing a 2000 bp alignment [74, 75]. In this first approach for the *trnL* intron, we analyzed a 450 bp alignment of 43 sequences from individuals belonging to six *C. brasiliense* populations. According to the above considerations, the number of populations is limited and the alignment extension should be increased. Both issues will be address in future analyses that will include unexplored *C. brasiliense* populations from Perú, Bolivia and northern Paraguay. In addition, other cytoplasmic and nuclear regions are being optimized for *C. brasiliense*, which will be useful in extending the alignments.

It is interesting that the Mexican and Paraguayan populations shared an inversion in the *trnL* intron. Although this must be confirmed through further analysis, the finding of this shared sequence variant led us to question about past dispersion routes of *C. brasiliense* from Mexico to Paraguay. At first, the Paraguayan populations appear to be remnants of Paranaense Forest Province that cover eastern Paraguay. In this sense, there are three records of *C. brasiliense* populations on the left side of the Paraguay River but no specimens have been registered along the banks of the Paraná River further North of San Ignacio, Argentina. One possible explanation for the absence of *C. brasiliense* along High Paraná River tract could be the result of the basaltic riverbed. The basaltic layer results in the river running fast and turbulent between steep banks making seed deposition difficult to occur. Sandstone outcrops in San Ignacio becomes the river broader for a short stretch letting sedimentation to occur. Once the river changes directions in the southwest plain, the current definitely turns calm and slow, allowing light seeds to be transported by flotation [76]. Thus, the species geographic discontinuity could be the consequence of a) the geomorphology characteristics and b) the temperature and rainfall patterns (Cardozo et al., *unpublished data*). All three could be limited factors for *C. brasiliense* colonization success; however, we can not exclude alternative colonization routes for the southernmost *C. brasiliense* populations analyzed in the present study. Another possible route to consider is the Paraguay River, which flows into the Paraná River, confluence that used to be further south and which has migrated towards the present Paraná River channel in its northward displacement. Could the Paraguay River have been a past southern communication bridge between the Paranaense and the Amazonic Biogeographic Subregions? This model seems to be supported by a phylogeographic study carried out in twenty-four *C. brasiliense* populations (23 from Brazil and one from Costa Rica) [77]. In the mentioned study the authors analyzed the chloroplastic intergenic region *trnH-psbA* and found a common haplotype within 14 populations from the Paranaense Subregion. Another nine haplotypes were identified in two and seven populations from the Amazonian and the Chaqueña Subregions, respectively, all of them restricted to a single population. These authors proposed that the Caatinga and Cerrado Provinces, that belongs to the Chaqueña Subregion, acts as a corridor joining the Amazonic and the Paranaense Forest Subregions. In the same way, the Chaco Province (also within the Chaqueña Subregion) located in the South of the Cerrado could also be consider a similar corridor. This proposition is in accordance with the fact that *C. brasiliense* is distributed all along the Neotropical Region. Interestingly, when we

compare the updated distribution map of *C. brasiliense* including Paraguayan and Argentinean location references with individual traces published [1], the major coincidence seems to be with the typical neotropical trace of the Hemipthera *Rhinacloa* (Figure 35, in Morrone, 2001). The analysis of samples derived from northern Paraguay, the western Brazilian States, Bolivia, Perú and northern and central populations that occupy the Paranaense Subregion would be crucial to elucidate the processes that originated the observed distribution patterns and to test the dispersion model we have proposed. If this was true and Paraguayan populations would have dispersed by the Paraguay River, our first hypothesis should be rejected at least in regard to the common origin.

The second assessment set out in our hypothesis affirms that the fragmentation process has provoked the differentiation of the analyzed *C. brasiliense* populations. This hypothesis has not been proved by the results obtained in the present work given that the Argentinean and Paraguayan populations all shared a largely distributed haplotype and represent a unique management unit (MU) in terms of conservation plans. However, the time passed since fragmentation events began could not be enough to allow mutations to accumulate and fix in the chloroplast genome of *C. brasiliense,* but would explain the high genetic differentiation observed between San Ignacio and Rincón Ombú populations carried out using arbitrary primers for nuclear genome, where the authors identified two MU [26]. Currently SSRs for nuclear *C. brasiliense* genome are being developed in order to analyze all known southern populations and as a whole, these data will increase the molecular information about these and other *C. brasiliense* populations. A deep population analysis using nuclear and cytoplasmic markers along with the ecological data that we are relieving about these hygrofile areas will largely contribute to our understanding of these austral forest patches and to drawing up management plans for these particular habitats conservation.

5. Conclusions

Three haplotypes were identified through which Argentinean, Paraguayan and Mexican populations could be differentiated. The haplotypic variants that have been found in the present work led us to reconsider and propose a model on *C. brasiliense'* geographic dispersion and colonization used in the past, both through the Paraná River and alternative routes.

Acknowledgements

References

[1] Morrone JJ (2001) Biogeografía de América Latina y el Caribe. M&T – Manuales & Tesis SEA, Zaragoza.

[2] Brieger FG (1969) Patterns of evolutionary and geographical distribution in neotropical orchids. Biol J Linn Soc 1(1-2):197-217.

[3] Müller P (1973) The Dispersal Centers of Terrestrial Vertebrates in the Neotropical Realm: A Study in the Evolution of the Neotropical Biota and Its Native Landscapes. The Hague: Dr. W. Junk.

[4] Leitão-Filho HF (1982) Aspectos taxonômicos das florestas do Estado de São Paulo. Silvicultura em São Paulo 16 A: 197-206.

[5] Rodrigues RR (2001) Florestas ciliares? Uma discussão nomenclatural das formações ciliares. In: Rodrigues RR, Leitão Filho HF (ed) Matas Ciliares: conservação e recuperação, 2nd edn. Editorial Universidad de São Paulo, São Paulo.

[6] Lobo PC, Joly CA (2001) Aspectos ecofisiológicos da vegetação de mata ciliar do sudeste do Brasil. In: Rodrigues RR, Leitão Filho HF(ed) Matas Ciliares: conservação e recuperação, 2nd ed. Universidad de São Paulo, São Paulo.

[7] de Souza A (2006) Estrutura genética de populações naturais de *Calophyllum brasiliense* Camb. na bacia de Alto Río Grande. Dissertation, Universidad Federal de Lavras.

[8] Fischer EA (1990) Distribuição de freqüência de clases de tamanho e estratégia reprodutiva de *Calophyllum brasiliense* Camb. (Guttiferae) em mata ciliar na Estação Ecológica Estadual Juréia-Itatins, São Paulo. Anais do II Simpósio de ecossistemas da costa sul e sudeste brasileira, ACIESP 71-1.

[9] Schiavini I (1992) Estrutura das comunidades de mata de galeria da Estação Ecológica do Panga (Uberlândia, MG). Dissertation. Universidad de Campinas, Brazil.

[10] Marques MCM (1994) Estudos auto-ecológicos do guanandi (*Calophyllum brasiliense* Camb. Clusiaceae) em mata ciliar no município de Brotas, SP. Dissertation. Universidad de Campinas, SP, Brazil.

[11] Kawaguici CB, Kageyama PY (2001) Diversidade genética de três grupos de indivíduos (adultos, jovens e plántulas) de *Calophyllum brasiliense* em uma população de mata de galeria. Sci For 59:131-143.

[12] Reitz R, Klein RM (1978) Projeto madeira de Santa Catarina. Sellowia 30: 28-30.

[13] Rodríguez ME, Cardozo A, Krauczuk E (2009) *Calophyllum brasiliense* (CLUSIACEAE) nuevo registro para la flora de la Argentina. Bol Soc Argent Bot 44(3-4): 361-366.

[14] Gasparotto Jr A, Benzan MA, Piloto IC et al (2005) Estudo fitoquímico da atividade moluscida do *Calophyllum brasiliense* Camb. (Clusiaceae). Química Nova 28: 575-578.

[15] Dharmaratne HRW, Tan GT, Marasinghe GP et al (2002) Inhibition of HIV-1 reverse transcriptase and HIV-1 replication by *Calophyllum* coumarins and xanthones. Planta Med 68: 86-87.

[16] Ito C, Itoigawa M, Mishina Y et al (2002) Chemical constituents of *Calophyllum brasiliense*: Structure elucidation of seven new xanthones and their cancer chemopreventive activity. J Natural Products 65: 267-27.

[17] Huerta-Reyes M, Basualdo MC, Lozada L et al (2004) HIV-1 Inhibition by extracts of Clusiaceae species from Mexico. Biol Pharm Bull 27: 916-920.

[18] Ito C, Itoigawa M, Mishina Y et al (2003) Chemical constituents of *Calophyllum brasiliense*. 2. Structure of three new coumarins and cancer chemopreventive activity of 4-substituted coumarins. J Natural Products 66(3): 368-371.

[19] Kimura S, Ito C, Jyoko N et al (2005) Inhibition of Leukemic Cell Growth by a Novel Anti- Cancer Drug (GUT – 70) from *Calophyllum brasiliense* that acts by induction of apoptosis. Int J Cancer 113: 158-165.

[20] Rea A, Tempone AG, Pinto EG et al (2014) Soulamarin Isolated from *Calophyllum brasiliense* (Clusiaceae) Induces Plasma Membrane Permeabilization of *Trypanosoma cruzi* and Mytochondrial Dysfunction. PLoS Negl Trop Dis 7(12): e2556. doi: 10.1371/journal.pntd. 0002556.

[21] Pires CTA, Brenzan MA, de Lima Socorro RB et al (2014) Anti-*Mycobacterium tuberculosis* activity and cytotoxicity of *Calophyllum brasiliense* Cambess. (Clusiaceae) Mem Inst Oswaldo Cruz 109(3): 324-329.

[22] Giraudo AR, Povedano H (2004) Avifauna de la región biogeográfica Paranaense o Atlántica Interior de Argentina: biodiversidad, estado del conocimiento y conservación. Miscelánea 12: 331–348.

[23] Galindo-Leal C, de Gusmão Câmara I (2005) Status do hotspot Mata Atlântica: uma síntese. In: Galindo-Leal C, de Gusmão Câmara I (ed) Mata Atlântica: Biodiversidade, Ameaças e Perspectivas.Center for Applied Biodiversity Science at Conservation International, Brasil.

[24] Galindo-Leal C, Jacobsen TM, Langhammer PF et al (2005) Estado dos hotspots: a dinâmica da perda de biodiversidade. In: Galindo-Leal C, de Gusmão Câmara I (ed) Mata Atlântica: Biodiversidade, Ameaças e Perspectivas. Center for Applied Biodiversity Science at Conservation International, Brasil.

[25] Moritz C (1994) Defining 'Evolutionarily Significant Units' for conservation. Tree 9(10): 373-375.

[26] Percuoco CB, Bich GA, Talavera Stéfani LN et al (2014) Assessment of genetic differentiation among relict populations of *Calophyllum brasiliense* Camb. (Calophyllaceae) from Northeast Argentina. JBES 5(6): 87-98.

[27] Ritland K, Clegg MT (1987) Evolutionary analysis of plant DNA sequences. Am Nat S74-S100.

[28] Chase MW, Soltis DE, Olmstead RG et al (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. Ann Missouri Bot Gard: 528-580.

[29] Olmstead RG, Palmer JD (1994) Chloroplast DNA systematics: a review of methods and data analysis. Am J Bot 81(9): 1205-1224.

[30] Drew BT, Ruhfel BR, Smith SA, Moore MJ, Briggs BG, Gitzendanner MA, Soltis PS, Soltis DE (2014) Another look at the root of the angiosperms reveals a familiar tale. Systematic Biology 63: 368–382.

[31] Petit RJ, Duminil J, Fineschi S et al (2005) Invited review: comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol Ecol 14 (3): 689-701.

[32] Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. J Mol Evol 28(1-2): 87-97.

[33] Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol Evol 16(3): 142-147.

[34] Ebert D, Peakall ROD (2009) Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. Mol Ecol Resour 9(3): 673-690.

[35] Huang S, Hwang S, Wang J et al (2004) Phylogeography of *Trochodendron aralioides* (Trochodendraceae) in Taiwan and its adjacent areas. J Biogeography 31: 1251–1259.

[36] Shaw J, Lickey EB, Beck JT et al (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot 92(1): 142-166.

[37] Shaw J, Lickey EB, Schilling EE et al (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot 94(3): 275-288.

[38] Huang H, Shi C, Liu Y, Mao S, Gao L (2014) Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evolutionary Biology 14: 151.

[39] Köppen W, Geiger R (1954) Klima der Erde (Climate of the Earth). Wall Map 1:16 Mill. Klett-Perthes, Gotha.

[40] Stange C, Prehn D, Arce-Johnson P (1998) Isolation of *Pinus radiate* genomic DNA suitable for RAPD analysis. Plant Mol Biol Rep 16: 1-8.

[41] Palmer JD (1987) Chloroplast DNA and biosystematic uses of chloroplast DNA variation. Am Nat 130: S6-S29.

[42] Shimada H, Sugiura M (1991) Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. Nucleic Acids Res 19: 983-995.

[43] Taberlet P, Gielly L, Pautou G et al (1991) Universal primers for amplifications of three non-coding regions of chloroplast DNA. Plant Mol Biol 17: 1105-1109.

[44] Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23(10): 1289-91.

[45] Untergrasser A, Cutcutache I, Koressaar T et al (2012) Primer3 - new capabilities and interfaces. Nucleic Acids Res 40(15): e115.

[46] Sambrook J, Fritsch E, Maniatis T (1989) Molecular clonning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.

[47] Zheng Z, Schwartz L, Wagner L et al (2000) A greedy algorithm for aligning DNA sequence. J Comput Biol 7(1-2): 203-214.

[48] Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41: 95-98.

[49] Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York.

[50] Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585-595.

[51] Fu Y (1997) Statistical tests of neutrality of mutations against population growth hitchhiking and background selection. Genetics 147(2): 915-925.

[52] Rozas J, Rozas R (1995) DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. CABIOS 11(6) 621-625.

[53] Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16: 37-48.

[54] Tamura K, Peterson D, Peterson N et al (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10): 2731-2739.

[55] Grivet D, Heinze B, Vendramin GG et al (2001) Genome walking with consensus primers: application to the large single copy region of chloroplast DNA. Mol Ecol Notes 1: 345-349.

[56] Shaw J, Shafer HL, Leonard OR et al (2014) Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. Am J Bot 101(11):1987-2004.

[57] Hwang SY, Lin TP, Ma CS et al (2003) Postglacial population growth of *Cunninghamia konishii* (Cupressaceae) inferred from phylogeographical and mismatch analysis of chloroplast DNA variation. Mol Ecol 12(10): 2689-2695.

[58] Lihová J, Kudoh H, Marhold K (2010) Genetic structure and phylogeography of a temperate-boreal herb, *Cardamine scutata* (Brassicaceae), in northeast Asia inferred from AFLPs and cpDNA haplotypes. Am J Bot 97(6): 1058-1070.

[59] Wheeler GL, Dorman HA, Buchanan A, Challagundla L, Wallace LE (2014) A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. Appl Plant Sci 2: 1400059.

[60] Bang SW, Chung S (2015) One size does not fit all: the risk of using amplicon size of chloroplast SSR marker for genetic relationship studies. Plant Cell Rep 34: 1681–1683.

[61] Petit RJ, Vendramin GG (2007) Plant phylogeography based on organelle genes: an introduction. In: Weiss S, Ferrand N (ed) Phylogeography of southern European refugia. Springer, Netherlands.

[62] Kelchner SA, Wendel JF (1996) Hairpins create minute inversions in non-coding regions of chloroplast DNA Curr Genet 30: 259-262.

[63] Byrne M, MacDonald B, Coates D (2002) Phylogeographical patterns in chloroplast DNA variation within the *Acacia acuminata* (Leguminosae: Mimosoideae) complex in Western Australia. J Evol Biol 15: 576–587.

[64] Borsch T, Quandt D (2009) Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. Plant Syst Evol 282(3-4): 169-199.

[65] Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. PloS one 6(5): e19254.

[66] Mustapha SB, Tamarzizt HB, Baraket G, Abdallah D, Salhi-Hannachi A (2015) Cytoplasmic polymorphism and evolutionary history of plum cultivars: Insights from chloroplast DNA sequence variation of trnL-trnF spacer and aggregated trnL intron & trnL-trnF spacer. Genetics and Molecular Research 14 (2): 3964-3979.

[67] Choulak S, Rhouma-Chatti S, Marzouk Z, Said K, Chatti N, Chatti K (2015) Chloroplast DNA analysis of Tunisian pistachio (*Pistacia vera* L.): Sequence variations of the intron trnL (UAA). Scientia Horticulturae 191: 57-64.

[68] Taberlet P, Coissac E, Pompanon F et al (2007) Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. Nucleic Acids Res 35: e14.

[69] Valentini A, Miquel C, Nawaz MA et al (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trnL* approach. Mol Ecol Resour 9: 1–60.

[70] Hale ML, Borland AM, Gustafsson MH (2004) Causes of Size Homoplasy Among Chloroplast Microsatellites in Closely Related Clusia Species. J Mol Evol 58: 182-190.

[71] Schaal BA, Leverich WJ, Rogstad SH (1991) A comparison of methods for assessing genetic variation in plant conservation biology. In: Falk DA, Holsinger KE (ed) Genetics and Conservation of Rare Plants. Oxford University Press, New York.

[72] Lu HP, Cai YW, Chen XY et al (2006) High RAPD but no cpDNA sequence variation in the endemic and endangered plant, *Heptacodium miconioides* Rehd. (Caprifoliaceae). Genetica 128(1-3): 409-417.

[73] Pons O, Petit RJ (1995) Estimation, variance and optimal sampling of gene diversity. Theor Appl Genet 90(3-4):462-470.

[74] Pluzhnikov A, Donnelly P (1996) Optimal sequencing strategies for surveying molecular genetic diversity. Genetics 144(3): 1247-1262.

[75] Felsenstein J (2006) Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? Mol Biol Evol 23(3): 691-700.

[76] Sanjurjo MD (1994) El arary, un árbol en extinción en el Paraguay. Revista Crítica 9: 51-55.

[77] Salgueiro F, Neri J, Alves-Ferreira M (2011) Phylogeographic patterns of *Calophyllum brasiliense* Camb. (Calophyllaceae) based on the psbA-trnH cpDNA locus. BMC Proceedings 5(Suppl 7): P17.